

Input file Flhl4273new; Output File Flhl4273tra

Sequence length 1743

TCCGGACTAGTTCCTAGACCGCTGCGGGCCGCCAGGCGCCCGGA	M S P E C A R A A	9
ATG TCC OCT GAA TGC GCG CGG GCA GCG		27
Å D A P L R S L E Q A N R R T R F P F F S		29
GCG GAC GCG CCC TTG CGC AGC CTG GAG CAA GCC AAC CGC ACC CGC TTT CCC TTC TTC TCC		87
D V K G D H R L V L A A V E T T V E V L		49
GAC GTC AAG GGC GAC CAC CGG CTG GTG CTG GCC GCG GTG GAG ACA ACC GTG CTG GTG CTC		147
I F A V S L L G N V C A L V L V A R R R R		69
ATC TTT GCA GTG TCG CTG CTG GGC AAC GTG TGC GCC CTG GTG CTG GTG GCG CGC CGA CGA		207
R R G A T A C L V L N L F C A D L L F I		89
GCG CGC GGC GCG ACT GCC TGC CTG GTA CTC AAC CTC TTC TGC GCG GAC CTG CTC TTC ATC		267
S A I P L V L A V R W T E A W L L G P V		109
AGC GCT ATC CCT CTG GTG CTG GCC GTG CGC TGG ACT GAG GCC TGG CTG CTG GGC CCC GTT		327
A C H L L F Y V M T L S G S V T I L T L		129
GCG TGC CAC CTG CTC TTC TAC GTG ATG ACC CTG AGC GGC AGC GTC ACC ATC CTC AGC CTG		387
A A V S L E R M V C I V H L Q R G V R G		149
GCG GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG CAC CTG CAG CGC GGC GTG CGG GGT		447
P G R R R A R A V L L A L I W G Y S A V A		169
CTC GGG CGG GCG GCG GCG GCA GTG CTG CTG GCG CTC ATC TGG GGC TAT TCG GCG GTC GCC		507
A L P L C V F F R V V P Q R L P G A D Q		189
GCT CTG OCT CTC TGC GTC TTC TTT OGA GTC GTC CCG CAA CGG CTC CCC GGC GCC GAC CAG		567
E I S I C T L I W P T I P G E I S W D V		209
GAA ATT TCG ATT TGC ACA CTG ATT TGG CCC ACC ATT CCT GGA GAG ATC TCG TGG GAT GTC		627
S F V T L N F L V P G L V I V I S Y S K		229
TCT TTT GTT ACT TTG AAC TTC TTG GTG CCA GGA CTG GTC ATT GTG ATC AGT TAC TCC AAA		687
I L Q I T K A S R K R L T V S L A Y S E		249
ATT TTA CAG ATC ACA AAG CCA TCA AGG AAG AGG CTC ACG GTA AGC CTG GCC TAC TCG GAG		747
S H Q I R V S Q Q D F R L F R T L F L L		269
AGC CAC CAG ATC CGC GTG TCC CAG CAG GAC TTC CGG CTC TTC CGC ACC CTC TTC CTC CTC		807
M V S F F I M W S P I I I T I L L I L I		289
ATG GTC TCC TTC TTC ATC ATG TGG AGC CCC ATC ATC ATC ACC ATC CTC CTC ATC CTG ATC		867
Q N F K Q D L V I W P S L F F W V V A F		309
CAG AAC TTC AAG CAA GAC CTG GTC ATC TGG CCG TCC CTC TTC TTC TGG GTG GTG GCC TTC		927

Fig-1

T F A N S A L N P I L Y N M T L C R N E 329  
ACA TTT GCT AAT TCA GCC CTA AAC CCC ATC CTC TAC AAC ATG ACA CTG TGC AGG AAT GAG 987

W K K I F C C F W F P E K G A I L T D T 349  
TGG AAG AAA ATT TTT TGC TGC TTC TGG TTC CCA GAA AAG GGA GCC ATT TTA ACA GAC ACA 1047

S V K R N D L S I I S G \* 362  
TCT GTC AAA AGA AAT GAC TTG TCG ATT ATT TCT GGC TAA 1086

TTTTTCTTTATAGCCGAGTTTCTCACACCTGGCGAGCTGGCATGCTTTTAAACAGAGTTCA<sup>TTT</sup>CCAGTACCTCCA

TCAGTGCACCTGCTTTAAGAAAAATGAACCTATGCAAAATGACATCCACAGCGTCGSTA<sup>TTA</sup>TTAAGGGGTGATCACCAA

GTTTCATAATATTTTCCCTTTATATAAANGGATTTGTGGCCAGGTGCAGTGGTTCATGCTGTAA<sup>TCC</sup>ACGAGTTTGGG

AGGCTGAGTGGGTGGATCACCTGAGGTGAGGAGTTCGAGACCAACCTGACCAACATGGTGAGACCCCGTCTCTACTA

AAAAATAAAAAAAAAAATTAGCTGGGAGTGGTGGTGGGCACCTGTAATCCTAGCTACTTGGGAGGCTGA<sup>ACC</sup>AGGAGAAT

CTCTTGAACCTGGGAGGCAGAGTTGCAGTGAGCGAGATCGTGCCATTGCACTCCAACAGGGCA<sup>AC</sup>AGAGTGAAAC

TCCATCTTAAAAAAAAAAAAAAAAAGATTGTTATGGGTTCCTTTTAAATGTGA<sup>ACT</sup>TTTTTTAGTGTGTTTGAATATG

ATCAAATTTAATAATATTTTATTTATGACTGTTCAGCAAAAAAAAAAAAAAAAAAGGGCGG

RTA01/2057957v1

FIG 1 (CONT)



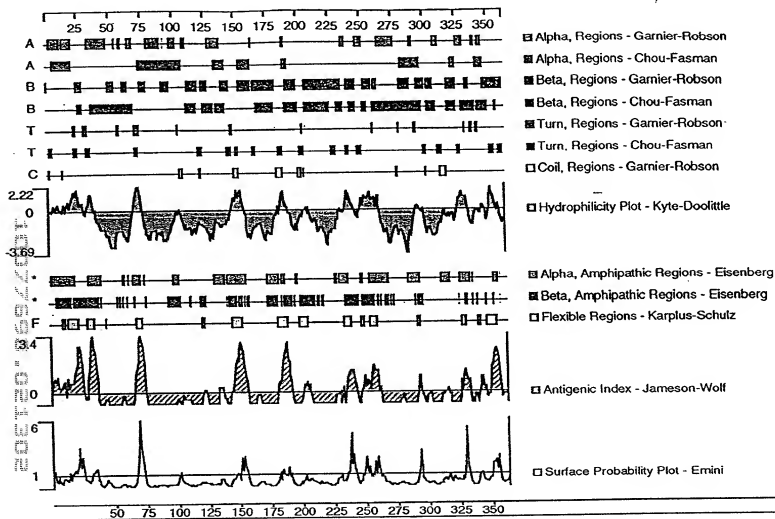


FIG 3

#

## Prosite Pattern Matches for Flh14273,

>PS00001|PDOC00001|ASN\_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTR 24

Query: 322 NMTL 325

>PS00004|PDOC00004|CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 239 KRLT 242

>PS00005|PDOC00005|PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239

Query: 350 SVK 352

>PS00006|PDOC00006|CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 256 SQD 259

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 57 GNVCAL 62

Query: 72 GATACL 77

Query: 343 GALLTD 348

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 150 PGR 153

>PS00028|PDOC00028|LEUCINE\_ZIPPER Leucine zipper pattern.

Query: 106 LGPVACHLLFYVMTLSG5VTIL 127

FIG 5

# Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
75	98	ins->out	4.0
113	134	out->ins	4.0
156	177	ins->out	4.9
209	227	out->ins	3.7
266	289	ins->out	6.5
297	321	out->ins	3.2

>Flh14273.  
 MSPECARAAGDAPLRSLQANRTREFFFSQVKGDRHLVLAAVETTVLVLPVAVSLGNYC  
 ALVLVARRRRRGATACLVNLFCADLLFISAIPVLAVRWTEAWLLGPFVACHLLFYVMTL  
 SSVTLTLAAVSLERMVCIHVLQGVGPGRRARAVLLALIWGYSAVAALPLCVFFRVV  
 PQLPGADQEISICTLIWPTIPGEISWDVSFVTLNPLVPLVIVISYSKILQITKASRRK  
 LTVSLAYSESHQIRVSQQDFRLFRFLFLMWVFFIMWSPILITILLIQNFQDLVIWP  
 SLFWVVAFTFANSALNPILYNMTCRNEWKKIPCCFWFPEKGAILTDTSVGRNDLSIIS

## Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
14	37	ins->out	4.0
52	73	out->ins	4.0
95	116	ins->out	4.9
148	166	out->ins	3.7
205	228	ins->out	6.5
236	260	out->ins	3.2

>Flh14273.\_mature  
 LVLVARRRRRGATACLVNLFCADLLFISAIPVLAVRWTEAWLLGPFVACHLLFYVMTLS  
 GSVTLTLAAVSLERMVCIHVLQGVGPGRRARAVLLALIWGYSAVAALPLCVFFRVV  
 QRLPGADQEISICTLIWPTIPGEISWDVSFVTLNPLVPLVIVISYSKILQITKASRRK  
 TVSLAYSESHQIRVSQQDFRLFRFLFLMWVFFIMWSPILITILLIQNFQDLVIWP  
 LFWVVAFTFANSALNPILYNMTCRNEWKKIPCCFWFPEKGAILTDTSVGRNDLSIIS

FIG-6

Input file 14273m; Output File 14273mtra  
Sequence length 1560

TTGCCAAGCTCAGCGTAAGCCTCTTCCACTGCAATCTCACAAGGGGTTTCATGGAGTGCTTCACACCATCAGTGACCA

CTCCAGACATTGTCCGGCTTTACCCGAATCTTCACAGCGAGTGGATGACCTCTTGACAGCCACGAGCGCGCGAGCTC

	M S P E C A O T T G	10
CGCCATCTTCCCGGACGCGTGGGCGGGCGCCCGGC	ATG TCC CCT GAG TGT GCA ACG ACG GGC	30
P G P S H T L D Q V N R T H F P F F S D	CCT GGT CCC TGG CAC ACC CTG GAC CAA GTC AAT CGC ACC CAC TTC CCT TTC TTC TGG GAT	30
V K G D H R L V L S V V E T T V L G L I	GTC AAG GGC GAC CAC CGG TTG GTG TTG AGC GTC GTG GAG ACC ACC GTT CTG GGA CTC AR	50
F V V S L L G N V C A L V L A R R R R R	TTT GTC GTC TCA CTG CTG GGC AAC GTG TGT GCT CTA GTG CTG GTG GCG GCG CGT CGG CGC	70
R G A S A S A S L V L N L F C A D L L F T S	OGT GGG GCG TCA GCC ACG CTG GTG CTC AAC CTC TTC TGC GGG GAT TTG CTC TTC ACC AGC	90
A I P L V L V V R W T E A W L L G P V V	GCC ATC CCT CTA GTG CTC GTC GTG GCG TGG ACT GAG GCC TGG CTG TTG GGG CCC GTC GTC	110
C H L L F Y V M T M S G S V T I L T L A	CTC CAC CTG CTC TTC TAC GTG ATG ACA ATG AGC GGC AGC GTG ACC ATC CTC ACA CTC GGC	130
A V S L E R M V C I V R L R R G L S G P	GCG GTC AGC CTG GAG CGC ATG GTG TGC ATC GTG GCG CTC CGG CGC GGC TTG AGC GGC CCG	150
G G R R T Q A A L L A F I W G Y S A L A A	GGG GCG CGG ACT CAG GCG GCA CTG CTG GCT TTC ATA TGG GGT TAC TCG GCG CTC GCC GCG	170
L P L Y I L F R V V P Q R L P G G D Q E	CTG CCC CTC TAC ATC TTG TTC CGC GTG GTC CCG CAG CGC CTT CCC GGC GGG GAC CAG GAA	190
I P I C T L D A W P N R I G E I S W D V F	ATT CCG ATT TGC ACA TTG GAT TGG CCC AAC CGC ATA GGA GAA ATC TCA TGG GAT GTG TTT	210
F E T L N F L V P G L V I V I S Y S K I	TTT GAG ACT TTG AAC TTC CTG GTG CCG GGA CTG GTC ATT GTG ATC AGT TAC TCC AAA ATT	230
L Q I T K A S R K R L T L S L A Y S E S	TTA CAG ATC ACG AAA GCA TCG CGG AAG AGG CTT ACG CTG AGC TTG GCA TAC TCT GAG AGC	250
H Q I R V S Q Q D Y R L F R T L F L L M	CAC CAG ATC CGA GTG TCC CAA GAC TAC CGA CTC TTC CGC AGC CTC TTC CTG CTC ATG	270
V S F F I M W S P I I I T I L L I L I Q	GTT TCC TTC TTC ATC ATG TGG AGT CCC ATC ATC ATC ACC ATC CTC CTC ATC TTG ATC CAA	290
N F R Q D L V I W P S L F F W V V A F T	AAC TTC CGG CAG GAC CTG GTC CAA CAA TCC CCA TCC CTT TTC TTC TGG GTG CTC TTC CTC TTC ACG	310
F A N S A L N P I L Y N M S L F R N E W	TTT GCC AAC TCT GCC CTA AAC CCC ATA CTG TAC AAC ATG TCG CTG TTC AGT AAG GAA TGG	330
R I F C C F F P E K G A I F T D T S	AGG AAG ATT TTT TGC TGC TTC TTT TTT CCA GAG AAG GGA GCC ATT TTT ACA GAT ACG TCT	350
V R R N D L S V I S S *	GTC AGG CGA AAT GAC TTG TCT GTT ATT TCC AGC TAA	362
		1082

CTAGCCTCTGGTGCCAGGTGAACACCGGTGTGATGTAAAGGGAGTTAACTTCAAGGAAAGCCACCAGTGCGCCCTTGC



TTTAAAAATACCCGACTTCCAACAGCAGGCATCTACGGAGCCAGCAAATTAAGGAATGATCGCTCAGTATAAAAATATT  
TTTCCTTAAAAGAACTTCTATGGGTCCTTTTGTGAACTTTTTTAAGTGTGTTTGTAAATATGATCTAGTTAATAAATT  
TTTATTATAACGTGTTCCTACAAAAAAAAAAAAAAAAAAAAA

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100

Query: 14273m,

Scores for sequence family classification (score includes all domains):

Sequence	Description	Score	E-value	N
7tm_1	PF00001 7 transmembrane receptor (rhodopsin	118.8	1e-35	1

Parsed for domains:

Sequence	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
7tm_1	1/1	57	321 ..	1	259 []	118.8	1e-36

Alignments of top-scoring domains:

7tm\_1: domain 1 of 1, from 57 to 321: score 118.8, E = 1e-36

```

*->GNILVilvrlrtkkrlrtpenifILNLAVADLLfltlppwalyylvg
GN+ ++++++R ++R ++ +LNL ADLLf+ +p++ ++ ++
GNVCALVLVAR-RRRRCASASLVNLFCADLLFTSAIPLVLVVR-WF 101

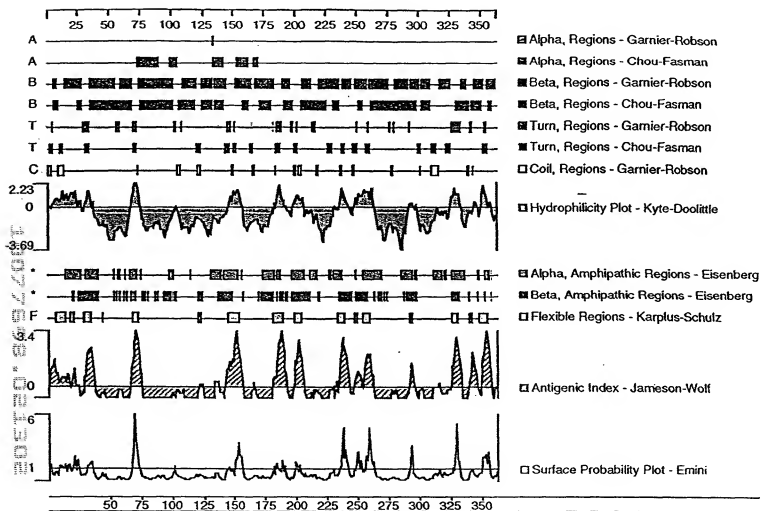
gsedWpfGsalCKlvtaldvnmmyaSiIlLtaISIDRYIAIvhlPrYrrr
e W++G+++C+L+ ++++++ + il+L+a S++R + IV L= +
14273m, 102 --EAWLLGPVVCHLLPYVMTKSGSVTILTLAAVSLERMVCIV-RLRRGLS 148

rtsprrA.kvillvWviallilsIPpllfswkvteegngtlnvrvvtCl
rr+++++++W ++l++lP ++++ v + ++G ++ +C+
14273m, 149 GP-GRRTQAALLAFIWGYSALAAFLYLFRVVPQRLPGGD--QEIPIC 195

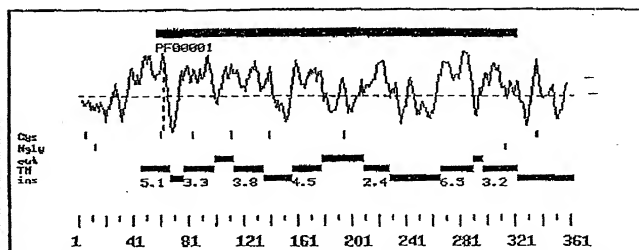
idfpeestasvstwlrsyvlstlvGFlllilvldvYtrtlrtlr....
+d+p++ + +s+ ++ ++ Fl+P lv++ Y+ Il + + +s+
14273m, 196 LDWPNRIG-----EISWDVFPETLNFLVPGLV--VISYSKILQITKasrk 239

.....kaaktllvvvvvfvICWlPyfivllldtl
+ + + +s+ + +s+ ++ +tl+vvv F++ W P i+ll +
14273m, 240 rltlslayseshqizvsgqyRLFRTLFLLMVSFFINWSPIIITLILLY 289

c.lsiimsstCelervlptallvllwLayvWscINPIIY<--
++ + + p ++++ + +s+N+INPi+Y
14273m, 290 QcFR-----QDLVINPFLFFWVVAITFANSALNPILY 321
  
```



# Analysis of 14273m, (362 aa)



>14273m, 1086 bases, 6943 checksum.  
 MSPECAQTGTGCPGSHTLQVNRTHFFFFSDVKGDERLVLVSVVETTVLGLIFVVSLLGNVC  
 ALVLVARRRRRGASASLVNLFCADLLFTSAIFLVLVVRWTEAWLLGPVVCHLLFYVMTM  
 SSVTLTLTAAVSLERMVCIIVRLRRGLSGPGRREQAALLAFINGYSALAALPLYILFRVV  
 FORLPGGDQRIPTCTLDWPNRIGEISWDVFFETLNLVPLVPLVIVISYSKILQITASRKE  
 LTLGLAYSESHQIRVSQQDYRLFTLLFLLMVSPFIMSPFIITILLILIQNFRQDCVIMP  
 SLFFWVVAFTFANSALNFIILYNHSLFRNWRKIFCCFFPERGAIFTDTSVRNDLSVIS  
 S\*

FIG 10

## Prosite Pattern Matches for 14273m,

>PS00001|PDOC00001|ASN\_GLYCOSYLATION N-glycosylation site.

Query: 21 NRTH 24

Query: 322 NMSL 325

>PS00002|PDOC00002|GLYCOSAMINOGLYCAN Glycosaminoglycan attachment site.

RU Additional rules:

RU There must be at least two acidic amino acids (Glu or Asp) from -2 to

RU -4 relative to the serine.

Query: 148 SGPG 151

>PS00004|PDOC00004|CAMP\_PHOSPHO\_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 239 KRLT 242

>PS00005|PDOC00005|PKC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.

Query: 237 SRK 239

Query: 350 SVR 352

>PS00006|PDOC00006|CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.

Query: 40 SVVE 43

Query: 256 SQDD 259

>PS00008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 57 GNVCL 62

Query: 72 GASASL 77

Query: 343 GAIFTD 348

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 150 PGRR 153

# Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
46	66	out->ins	5.1
77	98	ins->out	3.3
113	134	out->ins	3.8
156	177	ins->out	4.5
209	227	out->ins	2.4
266	289	ins->out	6.5
297	321	out->ins	3.2

>14273m,  
 MSPECAQTGPGPSHTLDQVNRTHFPFFSDVKGDHRLVLSVVETTVLGLIFVVSLLGNVC  
 ALVLVARRRRRGASASLVNLFCADLLFTSAIPVLVVRWTEAWLLGPVVCHLLFYVMCM  
 GSVITILTAAVSLERMVCIIVLRRLSGPGRRTQALLAFIWCYSALAAFLYILFRVVP  
 QRLPGGDQEIPICTLDWPNRIGETISWDVFETLNPLVPLVTVISYSKILQTKASRRL  
 TSLAYSESHQIRVQSODYRLERTFLLMVSPFIMWSPILITILLILIQNFRQDLVIWP  
 SLFPWWVAFTFANSALNPILYNMSLFRNEWKIFCCFFPEKGAIFTDSVRNDLSVIS  
 S

Transmembrane segments for presumed mature peptide

Start	End	Orient	Score
16	37	ins->out	3.3
52	73	out->ins	3.8
95	116	ins->out	4.5
148	166	out->ins	2.4
205	228	ins->out	6.5
236	260	out->ins	3.2

>14273m\_mature  
 LVVARRRRRGASASLVNLFCADLLFTSAIPVLVVRWTEAWLLGPVVCHLLFYVETMS  
 GSVITILTAAVSLERMVCIIVLRRLSGPGRRTQALLAFIWCYSALAAFLYILFRVVP  
 QRLPGGDQEIPICTLDWPNRIGETISWDVFETLNPLVPLVTVISYSKILQTKASRRL  
 TSLAYSESHQIRVQSODYRLERTFLLMVSPFIMWSPILITILLILIQNFRQDLVIWP  
 LFPWWVAFTFANSALNPILYNMSLFRNEWKIFCCFFPEKGAIFTDSVRNDLSVIS